International Journal of Engineering & Technology, 7 (4.10) (2018) 529-531

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Memoir on Model Selection Criterion between Two Nested and Non-Nested Stochastic Linear Regression Models

C. Narayana[1] , B. Mahaboob[2], B.Venkateswarlu[3]*, J. Ravi sankar[4]

[1]*Department of Mathematics, Sri Harsha institute of P.G Studies, Nellore.*
[2]*Department of Mathematics, K.L.E.F(Deemed to be University), Vaddeshwaram, Vijayawada, Andhra Pradesh.*
[3,4]*Department of Mathematics, VIT University, Vellore, Tamilnadu.*
[5]*Rtd. Professor, Department of Statistics, S.V. University, Tirupati, Andhra Pradesh.*
**Corresponding author E-mail: venkatesh.reddy@vit.ac.in*

Abstract

The main purpose of this paper is to discuss some applications of internally studentized residuals 9n the model selection criterion between two nested and non-nested stochastic linear regression models. Joseph et.al [1] formulated various proposals from a Bayesian decision-theoretic perspective regarding model selection Criterion. Oliver Francois et.al [2] proposed novel approaches to model selection based on predictive distributions and approximations of the deviance. Jerzy szroeter [3] in his paper depicted the development of statistical methods to test non-nested models including regressions, simultaneous equations. In particular new criteria for a model selection between two nested/ non-nested stochastic linear regression models have been suggested here.

Keywords: *Test statistic, OLS residual sum of squares, nested and non-nested stochastic linear regression model, internally studentized residuals, OLS estimator.*

1. Introduction

In In spite of the availability of highly innovative tools in Mathematics, the main tool of the Applied Mathematician remains the stochastic regression model in the form of either linear or nonlinear model. More importantly, mastery of the stochastic linear regression model is prerequisite to work with advanced mathematical and statistical tools because most advanced tools are generalizations of the stochastic linear regression model. The various inferential problems of stochastic modelling are considered to be essential to both theoretical and applied mathematicians and statisticians. The selection between alternative models is an important problem in stochastic modelling. Specification of the stochastic regression model is an important stage in any stochastic linear regression analysis. It includes specifying both the expectation function and the characteristics of the error. The various Miss-specification tests and testing general linear hypothesis in the stochastic linear regression models were studied by many mathematicians and statisticians. Most of these people have proposed their tests in stochastic linear regression models by using some inferential criteria.

2. Model Selection Criterion between Two Nested Stochastic Linear Regression Models Using Internally Studentized Residuals:

Consider two alternative nested stochastic linear regression models as

(i) $Y = X_1\beta_1 + \in_1$, such that $\in_1 \sim N(0, \sigma_1^2 I_n)$ and

(ii) $Y = X_1\beta_1 + X_2\beta_2 + \in_2$, such that $\in_2 \sim N(0, \sigma_2^2 I_n)$, where Y is $(n\times1)$ vector of observations on dependent variable; X_1 is $(n\times k_1)$ and X_2 is $(n\times k_2)$ data matrices of fixed known observations on independent variables and β_1 is $(k_1\times1)$ and β_2 is $(k_2\times1)$ vectors of unknown regression coefficients parameters , $\in_1$ and $\in_2$ are $(n\times1)$ vectors of errors and σ_1^2 and σ_2^2 are unknown error variances of $\in_1$ and $\in_2$ respectively; and $\in_1$, $\in_2$ follow Multivariate normal distributions with zero mean vectors and covariance matrices $\sigma_1^2 In$ and $\sigma_2^2 I_n$ respectively.

Model selection between these two nested stochastic linear regression models equivalently states the null hypothesis $H_0 : \beta_2 = 0$ against alternative hypothesis $H_1 : \beta_2 \neq 0$. For testing null hypothesis, one may fit these two models by Ordinary Least Squares (OLS) estimation procedure and then obtain the OLS residual sum of squares, and hence, internally Studentized residuals which are given by

OLS estimator of β_1: $\hat{\beta}_1 = (X_1'X_1)^{-1} X_1'Y$

Unbiased estimate of σ_1^2 : $\hat{\sigma}_1^2 = \frac{e_1'e_1}{n-k_1} = \frac{\sum_{i=1}^{n} e_{1i}^2}{n-k_1}$, where $e_1 = \left[Y - X_1\hat{\beta}_1\right]$

Internally Studentized Residuals:

$$q_{1i} = \frac{e_{1i}}{\hat{\sigma}_1\sqrt{1-h_{1ii}}}, \qquad i = 1,2,...,n$$

Where h_{1ii} = i^{th} diagonal element of Hat matrix, $H_1 = ((h_{1ij})) = X_1(X_1'X_1)^{-1}X_1'$

(i) Write $\beta = \begin{bmatrix}\beta_1 \\ \beta_2\end{bmatrix}$ and $X = [X_1 \quad X_2]$, one may have $Y = X\beta + \epsilon_2$, $\epsilon_2 \sim N(0, \sigma_2^2 In)$

OLS estimator of β : $\hat{\beta} = (X^1X)^{-1}X^1Y$

Unbiased estimator of $\hat{\sigma}_2^2 = \frac{e^1e}{n-k_1-k_2} = \frac{\sum_{i=1}^{n} e_i^2}{n-k_1-k_2}$

Where $e = [y - x\hat{\beta}]$

Internally studentized Residuals: $q_i = \frac{e_i}{\hat{\sigma}_2\sqrt{1-h_{ii}}}, i = 1,2,...,n$

Where h_{ii} = i^{th} diagonal element of Hat matrix H

$=((h_{ij})) = x(x^1x)^{-1}x^1$

Under the proposed model selection criterion, the test statistic for testing $H_0 : \beta_2 = 0$ is given by

$$F = \left[\frac{(q_1^1q_1 - q^1q)/k_2}{q^1q/(n-k_1-k_2)}\right] \sim F_{[K_2,(n-k_1-k_2)]}$$

Where $q_1^1q_1$: Internally studentized residual sum of squares obtained by using model

(i) q^1q : Internally studentized residual sum of squares obtained by using model

(ii) If F is not significant then Model (i) may be selected.

3. Model Selection Criterion between Two Nested Stochastic Linear Regression Models by Considering Omitted Independent Variables Are Unknown, Using Internally Studentized Residuals.

Consider two nested stochastic linear regression models under null and alternative hypotheses as $H_0 : Y_{n\times1} = X_{n\times k_1}\beta_{k_1\times1} + \epsilon_{n\times1}$ such that $\epsilon \sim N(0, \sigma^2 I_n)$.

Against $H_1 : Y_{n\times1} = X_{n\times k_1}\beta_{k_1\times1} + Z_{n\times k_2}\gamma_{k_2\times1} + \epsilon_{n\times1}$ such that $\epsilon \sim N(0, \sigma^2 I_n)$. Where Y is (nx1), β is (k_1x1), γ is (k_2x1), and ϵ is (nx1) vectors; X is (nxk$_1$) known and Z is (nxk$_2$) data matrix of omitted independent variables σ^2 is unknown scalar.

The problem of testing H_0 against H_1 can be equivalently state the null hypothesis as $H_0 : \gamma = 0$ against $H_1 : \gamma \neq 0$. Here, Z contains omitted independent variables which may cause inadequacy in the stochastic linear regression model. One may obtain Restricted and Unrestricted Internally Studentized residuals sum of squares by estimating two models under H_0 and H_1 respectively. The modified Wald test statistic for testing $H_0 : \gamma = 0$ given by

$$W^* = \frac{n(q_R^1q_R - q_{UR}^1q_{UR})}{q_{UR}^1q_{UR}} \overset{asy}{\sim} \chi^2_{k_2}$$

, Where $q_R^1q_R$: Restricted Internally Studentized Residual sum of squares obtained by using model under $H_0 : \gamma = 0$, $q_{UR}^1q_{UR}$: Unrestricted Internally studentized residual sum of squares obtained by using model under $H_1 : \gamma \neq 0$, K_2 : Number of restrictions on γ_2

If W^* is not significant then model under H_o may be selected as an adequate stochastic linear regression model.

4. Model Selection Criterion between Two Non-Nested Stochastic Linear Regression Models Using Internally Studentized Residuals

Consider two non-nested stochastic linear regression models under null and alternative hypotheses as

(i) $H_0 : Y = X\ \beta + \epsilon_1$ such that $\epsilon_1 \sim N(0, \sigma_1^2 I_n)$ against (ii) $H_1 : Y = Z\ \gamma + \epsilon_2$ such that $\epsilon_2 \sim N(0, \sigma_2^2 I_n)$

One may apply Ordinary Least Squares (OLS) estimation procedure and obtain the estimators of Y by using two models (i) and (ii) as $\hat{Y}_1$ and $\hat{Y}_2$ respectively.

ie : $\hat{Y}_1 = X\hat{\beta}_1$, where $\hat{\beta}_1 = (X^1X)^{-1}X^1Y$ and $\hat{Y}_2 = Z\hat{\gamma}$, where $\hat{\gamma} = (Z^1Z)^{-1}Z^1Y$ for testing the truth of H_o against H_1 consider an augmented stochastic linear regression model as

$Y = (1-\delta)\hat{Y}_1 + \delta\hat{Y}_2 + \epsilon_3$ Or $[Y - \hat{Y}_1] = \delta[\hat{Y}_2 - \hat{Y}_1] + \epsilon_3$ such that $\epsilon_3 \sim N(0, \sigma_3^2 I_n)$

$\Rightarrow Y^* = \delta X^* + \epsilon_3$, Which is a two variable linear model in deviation form.

Where $Y^* = [Y - \hat{Y}_2]$, is new dependent variable and $X^* = [\hat{Y}_2 - \hat{Y}_1]$ is new independent variable

Now, the OLS estimator of δ is given by $\hat{\delta} = \left[\frac{X^{*1}Y^*}{X^{*1}X^*}\right]$

For testing truth of H_o, one may test $H_{00} : \delta = 0$, by using Student's t-test statistic as

$$t = \frac{\hat{\delta}}{SE(\hat{\delta})} \sim t_{(n-1)}$$

where S.E. $(\hat{\delta})$ is standard error of $\hat{\delta}$.

For testing the truth of H_1 against H_0, one may consider an augmented stochastic linear regression model as $Y = \eta\hat{Y}_1 + (1-\eta)\hat{Y}_2 + \epsilon_4$, or $[Y - \hat{Y}_2] = \eta[\hat{Y}_1 - \hat{Y}_2] + \epsilon_4$, such that $\epsilon_4 \sim N(0, \sigma_4^2 I_n)$

$\Rightarrow Y^{k*} = \eta X^{**} + \epsilon_4$, Which is a two variable linear model in deviation form. Where $Y^{**} = [Y - \hat{Y}_2]$ is new dependent variable and $X^{**} = [\hat{Y}_1 - \hat{Y}_2]$ is new independent variable

Now, OLS estimator of η is given by $\hat{\eta} = \left[\frac{X^{**1}Y^{**}}{X^{**1}X^{**}}\right]$

For testing truth of H_1, one may test $H_{11} : \eta = 0$ by using Student's t-test statistic as $t = \frac{\hat{\eta}}{SE(\hat{\eta})} \sim t_{(n-1)}$. Compare the calculated values of t with their critical values and draw the inferences accordingly.

5. Conclusion

The above research study has brought out some new advance tools for analysing inferential aspects of stochastic linear regression models by using internally studentized residuals this work can be extended by developing advanced tools for analysing inferential aspects of stochastic non -linear regression models and random coefficients regression models by using different types of residuals other than studentized residuals. The research contribution. Made here could generate an immense interest in other research fellows to take up further research study in the coming years.

References

[1] Joseph B. Kadane, Nicol .A. Lazer (2003), "Methods and Criteria for model selection", Carnegie Mellon university Research, Department of Statistics.

[2] Neil H. Timm (2018), "Multivariate linear regression models, model selection, Fit association and prediction.

[3] V Olivier Francois and Guillaume Laval (2011), "Deviance information criteria for model selection in approximate Bayesian Computation", Statistical Applications in Genetics and Molecular biology, Vol.10, issue 1, Article 33.

[4] Ming ye, Philip D. Merger and Shlomo P. Neumam (2008), "On model selection Criterion in multi model analysis", Water resources research, Vol.44, Wo 3428.

[5] Jerzy Szroeter (1999), "Testing non-nested economic models", The current state of Economic Science, Pp: 223-253.

[6] Zucchini W. (2000), "An introduction to Model selection", Journal of Mathematical Psychology.

[7] Balasidda muni, P. et.ai. (2011), "Advanced Tools for Mathematical and Stochastic Modelling", Proceedings of the International Conference on Stochastic Modeling and Simulation, Allied Publishers.

[8] Byron J.T. Morgan, (2008), "Applied stochastic Modelling", CRC Press, 978-1-58488-666-2.

[9] Berry L. Nelson, (1995), "Stochastic Modelling, Analysis and Simulation", McGraw-Hill, 978-0070462137.

[10] Nelson, B.L. (1995), "Stochastic Modeling", McGraw-Hill, New York, 0-486-47770-3.

[11] Taylor, H.M. and Samuel karlin, (1998), "An Introduction to Stochastic Modelling", Academic Press, London, 978-0-12-684887-8.

[12] Nafeez Umar, S. and Balasiddamuni, P. (2013), "Statistical Inference on Model Specification in Econometrics", LAMBERT Academic Publishing, Germany.

[13] Ramana Murthy .B. et.al (2011), "A modified criterion for model selection", Proceedings of ICMS -2011, ISBN-978-81-8424-743-5.

[14] Rao .C.R. et.al (2001), "On model selection", Lecture notes-monographs series, infinite of Mathematical statistics, Vol.38, Pp: 1-64